Mixture of Expert/Imitator Networks: Scalable Semi-supervised Learning Framework

ID: 5069

Shun Kiyono[1] Jun Suzuki[1,2] Kentaro Inui[1,2]
[1]Tohoku University [2]RIKEN AIP

Motivation: "**Scalable**" Semi-supervised Learning (SSL) for NLP

Large-scale unlabeled data is available from Web

Common Crawl

Several terabytes
of data are available

Yet current SSL methods do not focus on scalability

- Often unlabeled data are used as input for complex DNN
  - e.g. Cross-view training [Clark+ 2018]
- Pretraining takes a month with 32 GPUs
  - ELMo [Peters+ 2018]

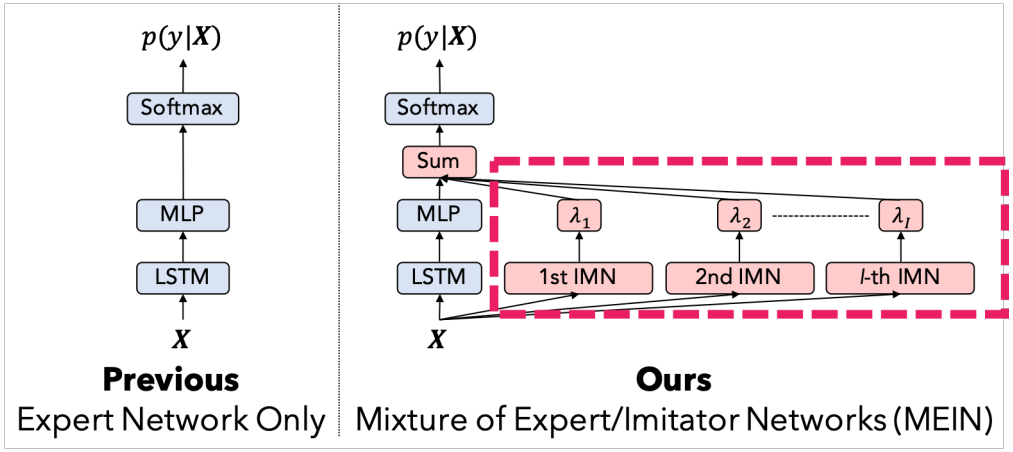
Several gigabytes
of data are used

Mixture of Expert/Imitator Networks: Scalable Semi-supervised Learning Framework

ID: 5069

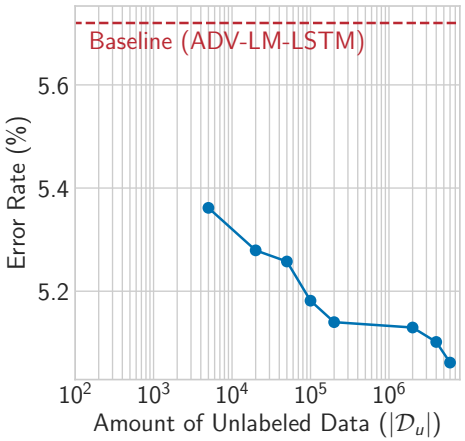
Shun Kiyono[1] Jun Suzuki[1,2] Kentaro Inui[1,2]
[1]Tohoku University [2]RIKEN AIP

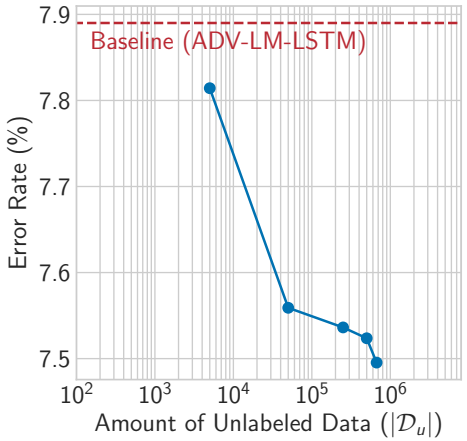
Proposed method consists of expert and imitator networks



We use unlabeled data to train Imitators (IMNs)

Our model has "more data, better performance" property





- More unlabeled data leads to better performance
- SOTA performance on text-classification task
- Proposed method is 8 times faster than current SSL method